基于 KECA 和 FWA-SVM 的间歇过程分时段故障诊断方法 *

蔡振宇，张　敏，包珊珊

(西南交通大学 机械工程学院，成都 610031)

摘　要：针对间歇过程的高度复杂性、强非线性、强时段性等特点，提出一种基于核熵成分分析（KECA）特征变量降维，利用烟花算法（FWA）优化支持向量机（SVM）参数的间歇过程分时段故障诊断方法。首先，通过多向核主元分析（MKPCA）进行在线故障监测，输出故障数据；其次，利用 K-means 分类方法将间歇过程划分为若干个子时段，对故障数据进行 KECA 特征变量处理，按熵值贡献率来确定选取主元的个数，深层提取特征信息；最后，在各子时段内分别构建 FWA 优化 SVM 参数故障诊断模型，将降维处理后的故障数据代入各自所属子时段 FWA-SVM 诊断模型内进行故障诊断。通过对青霉素仿真实验数据进行各种对比实验研究，验证了该方法的可行性与有效性。

关键词：间歇过程；核熵成分分析；烟花算法；支持向量机；K-means；青霉素仿真

中图分类号：TP277　　doi: 10.3969/j.issn.1001-3695.2017.12.0803

Time division fault diagnosis method based on KECA and FWA - SVM for batch process

Cai Zhenyu, Zhang Min, Bao Shanshan

(*School of Mechanical Engineering*, *Southwest Jiaotong University*, *Chengdu* 610031, *China*)

Abstract: Aiming at the high complexity, strong nonlinearity and strong time characteristics of intermittent process, this paper proposed a new method based on kernel entropy component analysis (KECA) to reduce the dimensionality of the KECA characteristic variables, and used the fireworks algorithm (FWA) to optimize the support vector machine (SVM) parameters for the intermittent process of division fault diagnosis method. Firstly, it carried out multi-directional kernel principal component analysis (MKPCA) for the on-line fault monitoring and output the fault data. Second, it used K-means method to divid the batch process into several sub-periods. It used KECA to reduce characteristic variable dimensionality according to the contribution rate of entropy to determine the number of selected elements and extracted feature information in depth. Finally, constructed FWA optimized SVM parameter fault diagnosis model in each sub-period, put the reduced dimension processed fault data into their own sub-period FWA-SVM diagnostic model for fault diagnosis. Through a variety of comparative experimental study based on penicillin simulation data, verified the feasibility and effectiveness of this method.

Key words: batch process; KECA; fireworks algorithm; support vector machine; K-means; penicillin simulation

0　引言

间歇过程广泛存在于现代生产生活当中，如食品、材料、化工和制药等[1]，因此实现高效故障监测与故障诊断至关重要。间歇过程生产周期短、操作过程重复性高、系统内部动态特性变化快，难以构造准确的数学模型[2]。主元分析(principal component analysis, PCA)[3]进行监测建模是一种线性建模方法，对于非线性系统如生物发酵往往不能保证其监测效果。针对非线性过程监测的建模问题，Scholkopf 等人[4]将核函数理论应用到统计过程监控中，提出了核主元分析（kernel principal component analysis, KPCA）。但 KPCA 算法不能直接在特征空间内进行数据重构与建立统计量监控图，难以直接应用于间歇过程在线监控。为此 Yoo 等人[5]提出了基于 KPCA 的非线性监测方法并应用于间歇过程在线监控，研究表明 KPCA 算法比 PCA 算法监控性能更优。多时段性研究[6]多存在于间歇过程故障监测方面，但间歇过程故障诊断也存在很强的时段性。

间歇过程变量维度过高，将变量进行一定的降维处理能提高故障诊断的稳定性与高效性。PCA[7,8]是应用广泛的经典数据降维方法之一，但对于非线性数据利用线性映射存在一定的局限性。KPCA 是 PCA 非线性的推广[9]，但数据处理与 PCA 类似，利用特征值大小来实现降维，降维效果存在一定的波动性。Jenssen[10]在 KPCA 的基础上提出核熵成分分析(kernel entropy

收稿日期：2017-12-11；**修回日期：**2018-01-29　**基金项目：**中央高校基本科研业务费专项资金资助项目（2682016CX031）；国家自然科学基金资助项目（51675450）

作者简介：蔡振宇（1992-），男，江西九江人，硕士研究生，主要研究方向为故障诊断、供应链管理（395966153@qq.com）；张敏（1986-），女，博士，讲师，主要研究方向为故障诊断、供应链管理；包珊珊（1993-），女，硕士研究生，主要研究方向为货位优化、供应链管理.

component analysis, KECA)算法用于数据降维。不同于 KPCA 与 PCA 算法，KECA[11]是计算 Renyi 熵来实现数据降维，在提取数据特征上表现出了其独特的优越性，比传统的 PCA、KPCA 更加稳定。

支持向量机 (support vector machine, SVM) [12]运用于故障诊断方面越来越成熟，但 SVM 通过核参数将特征向量映射到高维特征空间实现分类，参数的确定影响整个分类效果。利用传统的遗传算法(genetic algorithm, GA)、粒子群算法(particle swarm optimization, PSO)和交叉验证算法做 SVM 参数优化具有良好的效果，但是都存在一定缺陷。例如 GA 在计算过程中迭代收敛时间较长，PSO 则易受局部粒子最优影响，出现"早熟现象"等。何青等人[13]用果蝇优化算法优化 SVM 参数，但测试集量少，大量数据下还需进一步验证。烟花算法(firework algorithm, FWA)是 Tan 等人[14]在 2010 年提出的一种新型进化算法，具有很强的优化求解能力，并且能够在局部和全局搜索达到一个平衡效果，近年来逐渐受到研究者的关注。

基于此，本文提出一种新型、应用于间歇过程故障诊断的方法。首先采用多向核主元分析(multiway kernel principal component analysis, MKPCA)构建在线监测模型实现间歇过程故障监测，并输出故障数据；将间歇过程基于 K-means 方法划分为有限个子时段，标记故障数据所属子时段；通过 KECA 对故障数据进行变量降维，提取其有效信息；最后利用 FWA 优化 SVM 分类参数，在各子时段分别构建 FWA-SVM 诊断模型，将处理后的故障数据代入其对应的子时段模型内进行故障诊断。最后通过青霉素仿真实验数据进行仿真实验和各种对比实验研究验证了该方法的有效性。

1　MKPCA 间歇过程在线故障监测

1.1　数据预处理

间歇过程数据 $X(I\times J\times K)$，比连续过程数据多出一维批次元素，I 表示批量数，J 表示变量数，K 表示采样点数。将三维数据 $X(I\times J\times K)$沿批次方向展开得到 $X(I\times JK)$，每行是一个批次的所有数据，展开方式如图 1 所示。本文假设各个批次操作时间相同，在新构建的二维数据矩阵 $X(I\times JK)$中加入均值为零、方差为 0.01 的白噪声矩阵，去除噪声干扰，最后对合成的二维数据矩阵进行按列标准化处理。

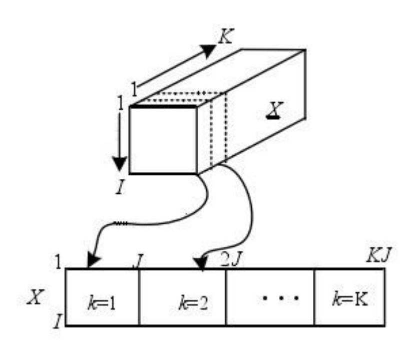


图 1　MKPCA 数据矩阵分解

1.2　在线故障监测

三维数据集 $X(I\times J\times K)$按上述沿批次展开后得到二维数据集 $X(I\times JK)$，再代入 MKPCA 模型内进行在线故障监测。首先利用正常工况下二维数据 $X(I\times JK)$进行离线分析计算统计量 Hotelling T^2 和平方预测误差(squared prediction error, SPE)的控制限，再进行在线监测。T^2 统计量通过表征模型内部变化来反映多变量变化情况，定义为

$$T^2 = t_i \lambda^{-1} t_i^T = x_i p \lambda^{-1} x_i^T p^T \tag{1}$$

T^2 统计量的控制限可利用 F 分布计算获得，如式（2）所示。

$$T_{\lim}^2 = \frac{p(n-1)}{n(n-p)} F(p, n-p) \tag{2}$$

其中：n 为样本数目；p 为主元个数。

平方预测误差 SPE 又可称为 Q 统计量法，表示 k 时刻的监测值对模型的偏离程度，是衡量模型外部数据变化的测度，定义为

$$SPE = \| \phi(x) - \phi_p(x) \|^2 = \sum_{i=1}^{n} t_i^2 - \sum_{i=1}^{p} t_i^2 \tag{3}$$

SPE 统计量在置信区间下的控制限通过正态分布确定：

$$SPE_{\lim} = \theta_1 [\frac{h_0 c_a \sqrt{2\theta_2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 + 1)}{\theta_1^2} + 1]^{1/h_0} \tag{4}$$

$$\theta_i = \sum_{j=a+1}^{n} \sigma_j^{2i} \tag{5}$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \tag{6}$$

其中：c_a 为与正态分布 $(1-\alpha)$ 分位点对应的标准差；α 为给定的显著水平。

当在线故障监测时，由于进行在线监测时所得数据只有从开始到监测时刻点的数据，但监测时需要整个批次数据，所以本文利用当前值对监测数据进行填充以满足整体批次数据，再计算监测数据采样点的统计量指标 T^2 和 SPE，与其相对应的控制限进行比较。当监测数据连续超过任一控制限大于等于 3 个采样时刻时，则判断当前有故障发生。

2　间歇过程的分时段故障诊断

2.1　K-means 聚类算法

间歇过程很少有论文进行分时段故障诊断研究，因此分时段故障诊断为本文一重点研究。间歇过程数据经过二维展开后，由于存在时段性，代入 K-means 算法内进行时段划分。K-means 算法的原理是先随机产生 L 个初始位置点作为 L 个簇的初始中心点，将邻近的点分到最近的簇，然后计算各个簇的质心，重新确定新的质心。如此不断地进行迭代，直到质心的移动范围满足所需要求或者迭代要求。

假设初始聚类中心 $Z=\{z_1, z_2, \ldots z_L\}$ 及聚类数据 $X=\{x_1, x_2, \ldots x_n\}$，则 K-means 算法具体操作步骤分为以下 5 步：

a）从 n 个数据中随机选取 L 个对象作为初始聚类中心。

b）计算每个数据对象 $x_i \in X$ 到 L 个聚类对象 $z_j(j=1,2,...L)$ 的距离，本文采用马氏距离。

c）聚类完成后重新计算该聚类中距离每一个点的最近的中心点。

d）若与上一次计算出的聚类中心相同，说明聚类结果成立，则转到 e）；若与上一次计算出来的聚类中心不同，说明聚类还未完成，则转到 b）重新进行计算。若满足迭代要求，则转到 e）。

e）结束并输出聚类结果。

2.2　故障特征提取与降维处理

当在线监测判断当前批次数据有故障发生时，由于间歇过程是复杂生产过程，往往会使故障变量随后传递，掩盖后面变量信息，所以只提取发生故障之后的若干采样点作为故障数据，使获得的故障数据更具有代表性，提高故障诊断率。间歇过程数据具有非线性和变量维数较高的特性，本文采用 KECA 算法对故障数据进行特征降维，提取有效信息，从而实现高效故障诊断。

KECA 是一种不同于传统“谱方法”的非线性降维算法，能够将数据降维前后的 Renyi 熵损失最小化。设某一概率系统中有 n 个事件或数据集 $(X_1, X_2, ...X_n)$，第 i 个事件 X_i 产生的概率为 $P_i(i=1,2...n)$，则其数据的 Renyi 熵指标可以表示为

$$H(p) = -\lg\int p^2(x)dx \tag{7}$$

对数函数是单调函数，因此可以只对积分一部分进行单独定义：

$$V(p) = \int p^2(x)dx \tag{8}$$

需要对 $V(p)$进行估计计算，才可以求出 Renyi 熵。文中引入 Parzen 概率密度算子[15]，根据高斯函数卷积理论并结合函数的单调性，化简得到

$$\hat{V}(p) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K_{\sqrt{2}\sigma}(x_i, x_j) = \frac{1}{n^2}I^TKI \tag{9}$$

$$K_{i,j} = K_{\sqrt{2}\sigma}(x_i, x_j) \tag{10}$$

其中：I 为每个元素均为 1 的 N 维向量；K 为样本核矩阵。

通过式(11)实现了 Renyi 熵的核矩阵表达，对核矩阵进行分解，表示为

$$K = ED_\lambda E^T \tag{11}$$

其中：D_λ 为特征值 $\lambda_1,...\lambda_n$ 组成的对角矩阵；E 为对应的特征向量 $e_1,...e_n$ 组成的矩阵。

因此，式（9）进一步可表示为

$$\hat{V}(p) = \frac{1}{n^2}(\sum_{i=1}^{n}\sqrt{\lambda_i}e_i^TI)^2 \tag{12}$$

式（12）表明每个 K 的特征值和特征向量对熵值的估计都有自己的贡献，但贡献大小并不一致。

2.3　多分类支持向量机

支持向量机实现多故障诊断时是通过构造多个二分类器组合实现多分类功能，并利用核函数来实现线性不可分向线性可分的转换。研究表明径向基核函数在 SVM 中表现出良好的泛化能力，其形式如式(13)所示。

$$K(y_i, z_j) = \exp(-\frac{|y_i - z_j|^2}{\sigma^2}) = \exp(-\gamma|y_i - z_j|^2) \tag{13}$$

将输入向量从原来的空间映射到高维特征空间 H，并在该特征空间 H 内建立优化超平面。分类线方程[16]如下：

$$\omega \cdot x + b = 0 \tag{14}$$

$$y_i[\omega \cdot x_i + b] \ge 1 \qquad i = 1,2,...l \tag{15}$$

其中：ω 为权值；$\boldsymbol{x}$ 为输入向量；b 为阈值；l 为向量的个数。

根据 Karush—Kuhn—Tucker，优化各个系数得到最优分类函数：

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{i} y_i a_i \sigma(x, x_i) + b) \tag{16}$$

适应度函数则为

$$f(\sigma^2, C) = \frac{1}{R(\sigma^2, C)} \tag{17}$$

由式（17）可知，进行 SVM 分类模型构建时，性能的关键因素在于惩罚参数 C 和核函数 σ 的选取，本文通过选用烟花算法对这两类参数进行寻优。

2.4　新型群体智能算法—烟花算法

烟花算法是将每个烟花都当做一个解空间中的可行解，通过爆炸产生不同的烟花点作为在可行域内进行全局搜索的可行解。通过每个烟花的适应度值来确定爆炸半径和爆炸数。适应度值越小的点，爆炸范围越小，爆炸数越多，适应度值大的则相反。烟花算法核心包括爆炸算子、变异操作、映射规则、选择策略四部分[17]。FWA 寻优收敛具体步骤如下：

a）在解空间内随机初始 N 个位 x_i，有 N 个烟花；

b）计算每个烟花的适应度值，与它们的爆炸半径 Rad_i 和爆炸火花个数 SS_i 为

$$Rad_i = RC \times \frac{f(x_i) - y_{\min} + \varepsilon}{\sum_{i=1}^{N}(f(x_i) - y_{\min}) + \varepsilon} \tag{18}$$

$$SS_i = H \times \frac{y_{\max} - f(x_i) + \varepsilon}{\sum_{i=1}^{N}(y_{\max} - f(x_i)) + \varepsilon} \tag{19}$$

其中：$y_{\min} = \min(f(x_i))(i=1,2,...N)$ 是当前迭代中的最优值，也为最小值；而式 $y_{\max} = \max(f(x_i))(i=1,2,...N)$ 为当前迭代中的最劣值，也为最大值；RC 和 H 分别表示用来调整爆炸半径的大小和爆炸火花数的大小；ε 表示机器最小量，避免出现零操作。

为了避免爆炸火花优或劣的适应度值产生过多或过少，文献[14]对火花个数作出如下的限制：

$$S_i=\begin{cases} round(a*H), S_i<aH \\ round(b*H), S_i>bH, a<b<1 \\ round(S_i), \quad 其他 \end{cases} \tag{20}$$

c）产生爆炸火花，集合 DC 具有 z 个维度，$z=round(D\times rand(0,1))$，在 DC 中的每个维度 k 下进行爆炸操作后，再经过越界处理将 ex_{ik} 保存到火花种群中。

$$h=Rad_i\times rand(-1,1) \tag{21}$$

$$ex_{ik}=x_{ik}+h \tag{22}$$

其中：h 为偏移量；x_{ik} 为第 i 个烟花的第 k 维；ex_{ik} 为 x_{ik} 爆炸操作后的火花。

d）进行高斯变异操作，每个维度通过式（23）进行高斯变异后，再经过越界处理保存到高斯种群当中。

$$mx_{ik}=x_{ik}\times e \tag{23}$$

e）选择操作，在所有得到的种群中挑选最好的一个，另外 N-1 个则进行轮盘赌法进行选择。

$$p(x_i)=\frac{R(x_i)}{\sum_{x_j\in K}R(x_i)} \tag{24}$$

$$R(x_i)=\sum_{x_j\in K}d(x_i-x_j)=\sum_{x_j\in K}\|x_i-x_j\| \tag{25}$$

f）判断是否满足终止迭代条件，若满足输出最优目标值，不满足就继续迭代。

2.5　基于 FWA-SVM 故障诊断模型构建

利用烟花算法对 SVM 的参数进行寻优，其步骤如下所示：

a)设定 SVM 参数范围。确定式（17）的两个参数所选的范围。

b)烟花种群初始化。设定烟花种群个数、子代数、高斯算子、最大迭代数等参数。

c)故障数据 KECA 降维。将每个子时段故障数据进行 KECA 处理，按照熵值大小获得主元结构实现数据降维，以便于代入故障诊断模型内进行高效诊断。

d)FWA-SVM 模型训练。将处理后的故障数据代入各自的子时段 FWA-SVM 模型内进行训练。

e)输出最优参数。当迭代次数满足终止迭代条件后各子时段模型输出最优参数构建最优诊断模型。

基于 FWA-SVM 故障诊断模型构造流程如图 2 所示：

2.6　基于 KECA 和 FWA-SVM 的间歇过程分时段故障诊断

本文提出的间歇过程分时段故障诊断分为离线各分类器的构建和在线故障诊断。其中离线分类器计算步骤如下：

a)将间歇过程划分各时段，通过 K-means 划分算法划分为若干个子时段。

b)收集各子时段所有故障的 M 个批次数据，定义故障类别后再进行数据预处理。

c)在预处理后的数据中提取发生故障时刻之后若干个采样点数据作为训练集数据。

d)得到的训练集数据进行 KECA 处理，按照熵值信息大小选择主元结构进行降维。

e)将各时段的训练集数据代入所属子时段 FWA-SVM 诊断模型内进行训练，构建最优诊断模型。

进一步实现间歇过程的在线故障诊断，其步骤如下：

a)通过正常工况下的若干批次构造 MKPCA 监测模型，计算 T2 和 SPE 统计量 99%的控制限。

b)将采样点利用当前值进行填充，对完整批次数据做数据预处理后计算两个统计量 T2 和 SPE，监测判断采样点是否连续三次超过各自的控制限，若超出则有故障，否则正常。

c)提取故障发生时刻后若干个点采样时刻的故障数据，进行 KECA 降维处理后再送入训练好的 FWA-SVM 诊断模型内进行故障诊断。

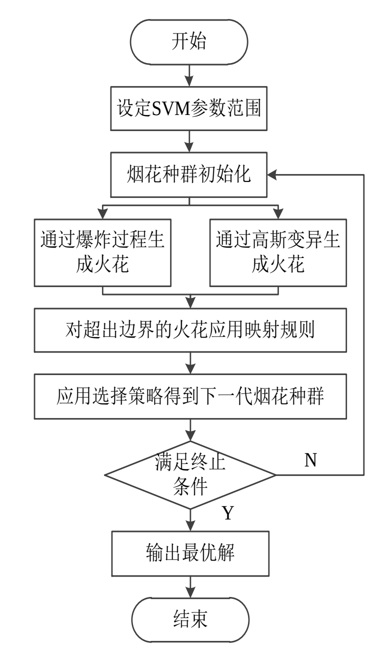


图 2　FWA-SVM 故障诊断模型构造流程

3　实验仿真

青霉素发酵过程是一种典型的间歇过程，本文采用青霉素生产仿真软件 Pensim2.0[18]生成实验数据。为实现准确的在线故障监测，利用 Pensim 模型模拟发酵时间为 400 h、采样间隔为 0.5 h 的 140 批次正常数据，青霉素仿真数据共有 18 维变量数据。监测时为简便而高效只选取通风率、搅拌器功率等 10 个影响青霉素发酵过程的变量，避免过程变量过多，使监测效果不明显。选取变量如表 1 所示。

间歇过程具有时段特性，将初始三维数据矩阵 $X(I\times J\times K)$按采样点展开为二维矩阵数据 $X(IJ\times K)$，并对其进行标准化处理，利用 K-means 划分算法划分间歇过程，参考青霉素发酵过程中各个变量的轨迹图[19]选取 K-means 参数 k 为 4，实验若干次后选取端点出现频率最高的作为时段分隔点，阶段划分结果为 0~49 h、49~120 h、120~255 h、255~400 h。利用 Pensim 模型模拟发酵时间为 400 h、采样间隔为 0.5 h，并且在每个子时段都生成故障 1、2、3 各 15 组共 180 批次的故障数据，如表 2 所

示。故障 1、2、3 依次为通风率故障、搅拌功率故障和底物流加速率故障。

表 1　在线监测建模所选变量

参数属性	变量	单位	取值范围
初始条件设置	底物浓度	G/L	14-18
	溶解氧浓度	G/L	1-1.2
	二氧化碳浓度	G/L	0.5-1
	pH 值		4.5-5.5
	温度	K	295-301
控制参数设置	通风率	L/H	8-9
	搅拌功率	W	29-31
	底物流加速率	L/H	0.039-0.045
	底物流温度	K	295-296
	反应器温度	K	297-298

表 2　3 类不同故障的所属时段与组数

时段/h	故障 1	故障 2	故障 3
0-49	15 组	15 组	15 组
49-120	15 组	15 组	15 组
120-255	15 组	15 组	15 组
255-400	15 组	15 组	15 组

3.1　在线监测仿真

先将 140 组正常工况批次的三维数据进行数据预处理，包括沿批次展开、添加噪声信号和进行数据标准化；再取前 40 组正常批次数据计算 T^2 和 SPE 统计量的 99%的控制限。为研究 MKPCA 运用在本论文中的监测效果，取后 100 组正常批次数据用来研究报错率。另外，将所有故障数据进行正常批次相同的数据预处理，包括选取 10 个相关变量、添加白噪声等处理。进行在线监测时，由于需要一个整体批次信息才能实现监测，而在线监测时只有从开始到当前监测采样的数据信息，本文采用当前值填充监测数据以满足整体批次数据要求实现监测。在线监测实验结果分别为监测报错率 5%、监测诊断率 97.78%。

从检测的诊断效果和报错率上可以说明 MKPCA 在间歇过程上进行监测效果较好。虽然报错率 5%相对较高，但对比于诊断率 97.78%来说还是相当可观，因为对于大多数间歇过程，达到诊断率更高的实用效果更好。因此 MKPCA 在一定程度上还是可以准确地进行在线故障监测，准确监测故障并输出故障数据集及其各所属子时段，为接下来的故障诊断做好充足、准确的数据准备。

3.2　故障诊断仿真实验研究

为提高监测效率，只提取监测数据一部分相关变量，但对于故障诊断，变量数据大多都有相关性。因此监测出某批次连续三次超出任一控制限的故障数据后取连续 20 个采样点，还原该采样点所有变量和其各所属时段信息。本文采用的青霉素仿真实验数据过程变量有 18 维变量，去除第一维采样点数变量后，对剩下的 17 维进行数据分析与故障诊断。利用各时段已知三类故障作为训练集训练诊断模型，对监测出故障的数据集进行故障诊断。

3.2.1 数据降维研究对比

故障诊断时需要对故障数据集做数据预处理，本文为研究 KECA 数据降维对间歇过程故障诊断的有效性，对已知故障数据集和监测输出的故障集不进行数据降维处理、KPCA 降维处理和 PCA 降维处理与进行 KECA 降维处理进行对比实验。为单独研究 KECA 数据降维效果，不添加其他方面因素影响处理效果，本文利用寻优效果良好的遗传算法优化 SVM 参数构造的模型进行验证对比，数据采用子时段 255~400 h 故障数据，结果是重复 5 次的平均值，如图 3 和表 3 所示。

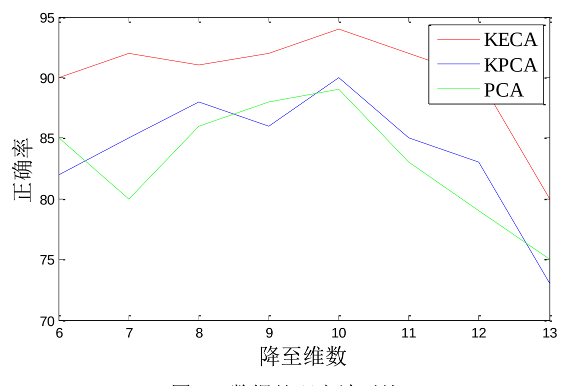


图 3　数据处理方法对比

从图 3 可以看出，当两者数据处理方法都降至 10 个维度时正确率最高，此时 KPCA、PCA 贡献率分别为 97%、90%，而 KECA 则超过了 99%的熵值信息。从表 3 中可以看出，无论是训练集正确率还是测试集正确率，KECA 方法都明显高于 KPCA 与 PCA 方法。不进行数据降维的诊断效果远弱于前三者，即使训练集正确率很高，但正确率低，在训练时出现过拟合现象，导致测试效果差。因此将数据作正确的处理对诊断结果来说很有必要，可以实现故障的高效诊断。

表 3　三种数据处理方法对比研究

算法	降至维数	训练集正确率	正确率
无	17	98.5%	78.833%
PCA	10	90.333%	86.655%
KPCA	10	92.667%	89.333%
KECA	10	97.667%	94%

3.2.2 优化算法研究对比

将各子时段已知故障数据进行 KECA 数据降维后，生成各子时段的训练集训练模型。构造基于 PSO 优化 SVM 模型、GA 优化 SVM 模型和交叉验证算法优化 SVM 模型，与本文提出的基于烟花算法优化 SVM 模型进行对比，分别记为 PSO-SVM、GA-SVM、CV-SVM、FWA-SVM.各算法初始参数如表 4 所示。利用时段 255~400 h 故障数据进行故障诊断，结果如表 5 所示。

从表 5 可以看出，烟花算法优化 SVM 模型时间上远远优

于 GA 和 PSO，略快于交叉验证算法，主要在于算法本身构造上的不同：烟花算法通过分布式信息共享，根据分布在不同区域烟花的适应度值决定爆炸强度大小和辐射范围；PSO 是单项流动，搜索迭代过程是跟随当代最优解；GA 通过选择、交叉、变异等操作，一代一代寻找最优；CV 则是按照网络划分进行层层搜索。再看 FWA-SVM 模型正确率远远优越于 PSO，略高于 GA 和 CV，从寻优效果来说是最好的，验证了该方法的有效性。综上，从迭代收敛时间和寻优正确率来看，利用烟花算法优化 SVM 参数应用于间歇过程来说是最适合，效果最佳。

表 4　各算法初始参数设定表

算法	参数	对应值或区间
PSO	粒子种群大小	20
	最大迭代数	400
	惯性权重	[0.3,0.9]
	学习因子 1、2	1.2、1.5
GA	种群数大小	20
	最大迭代数	400
	交叉概率	0.4
	变异概率	0.01
CV	V 参数大小	3
	c 参数范围	$[2^{-4},2^{7}]$
	g 参数范围	$[2^{-6},2^{10}]$
	c 步进大小	1
	g 步进大小	1
FWA	烟花种群大小	20
	最大迭代数	400
	爆炸半径幅度	40
	爆炸火花数范围	[1,40]
	高斯变异火花数	5

表 5　四种模型分类结果对比

模型	迭代时间	正确率
PSO-SVM 模型	117.56s	86.25%
GA-SVM 模型	87.725s	94.26%
CV-SVM 模型	15.21s	94.33%
FWA-SVM 模型	13.36s	95%

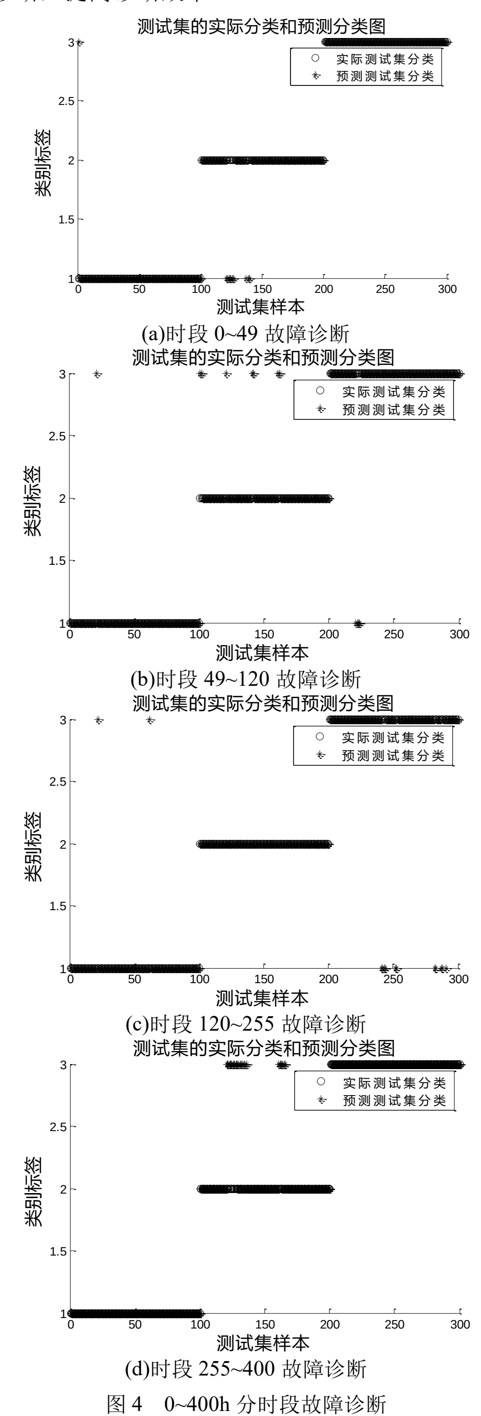
3.2.3 分时段与整体时段研究对比

间歇过程存在时段性，同一故障不同的时段会表现出不同的特征状态。本文通过 K-means 划分算法将青霉素发酵过程合理整合成四个子时段，分别为 0~49 h、49~120 h、120~255 h、255~400 h，每个子时段都利用 FWA-SVM 模型进行故障诊断。作为对比的整体时段 0~400 h 构建一个 FWA-SVM 模型。将监测出故障的数据进行数据处理后诊断故障，结果对比如表 6 和图 4、5 所示。

表 6　阶段效果对比

时段	5 次平均诊断率	平均诊断率
0-49h	97.33%	93.99%
49-120h	91.33%	
120-255h	92.33%	
255-400h	95%	
0-400h	88.75%	88.75%

分时段进行故障诊断的平均诊断率是 93.99%，比整体诊断率高 5%左右，并且所有时段诊断率都高于整体诊断率，这充分说明了间歇过程的时段性。因此，在进行间歇过程的故障诊断时应该先从阶段性分析，将过程合理分为若干个子时段，再进行故障诊断，提高诊断效率。



(a)时段 0~49 故障诊断
(b)时段 49~120 故障诊断
(c)时段 120~255 故障诊断
(d)时段 255~400 故障诊断

图 4　0~400h 分时段故障诊断

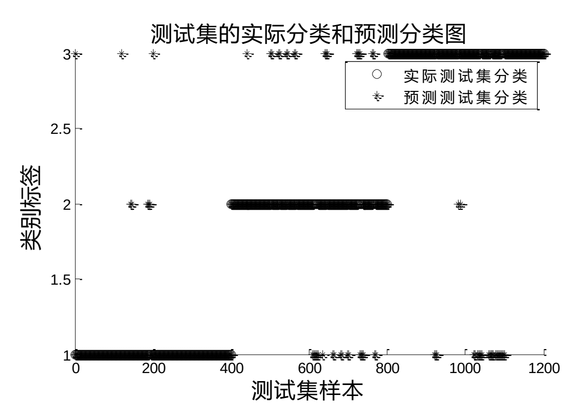


图 5 0~400 h 整体故障诊断

3.2.4 与文献研究对比

为说明本文提出的算法运用在间歇过程中的实效性，本文按照文献[20]提取数据方法取发生故障后的 10 个连续数据作为测试数据，在时段 297~400 h 后每类故障增加 5 批次，以满足在 49~400 三个时段内每类故障都有 50 批次，并且故障幅度都在[-5%;10%]之间。对故障 1、2、3 在起始时间 50~380 h 内按照本文方法和文献方法进行诊断来研究对比。结果如表 7 所示，数据结果为实验 10 次的平均值。

表 7 两种方法对故障诊断研究对比

故障类型	文献[18]识别率	本文识别率
故障 1	100%	99.55%
故障 2	94%	94.33%
故障 3	92%	96.44%
总体	95.33%	96.77%

与文献[20]对比，在故障 3 识别效率有明显的改进，故障 2 略高，故障 1 诊断率略差 0.45%，但对于总体识别效率具有较大改进。从两者故障数据选择上，本文选择将所有间歇过程变量都选入，并作 KECA 降维处理，很大程度上保证了故障数据的完整性，也能提高诊断效率；两者故障诊断模型构造也不一样，本文选择用 FWA 优化 SVM 参数，可很大程度上进行模型优化，使故障诊断率更高，而文献[20]采用 LSSVM 诊断模型，选择核函数时参数并未优化，因此效果不佳；另外本文提出进行分时段故障诊断，在不同时段内构造诊断模型可以很大程度上提高诊断效率。综上，本文提出的基于 KECA 特征提取与 FWA-SVM 模式识别间歇过程分时段故障诊断方法相比文献[20]有较好的改进，验证了该方法的可行性与有效性。

4 结束语

本文通过对间歇故障研究，利用 MKPCA 监测间歇过程，将监测出故障采样点的数据集进行 KECA 数据降维，通过 K-means 对间歇过程划分若干个子时段，在每个子时段内构建 FWA-SVM 诊断模型，再将经过 KECA 处理后的数据集代入各自模型内进行故障诊断。利用青霉素间歇过程实验仿真数据进行实验仿真，表明 KECA 数据降维能对间歇过程数据实现很好的数据处理；新型的烟花算法在优化 SVM 参数时比 GA、PSO 和 CV 迭代收敛速度快、识别率更高；同时讨论了间歇过程时段性强，采用分时段研究能够更好的实现各故障的诊断。

参考文献：

[1] Zhang C, Li Y, Study on the fault-detection method in batch process based on statistical pattern analysis [J]. Chinese Journal of Scientific Instrument, 2013, 34 (9): 2013-2110.

[2] Kerkhof P V D, Gins G, Vanlaer J, *et al.* Dynamic model-based fault diagnosis for (bio) chemical batch process [J]. Computers & Chemical Engineering, 2012, 40 (10): 12-21

[3] 张珂，宋文丽，石怀涛. 基于改进核主元分析的故障检测方法研究 [J]. 控制工程, 2017, 24 (2): 418-424.

[4] Scholkopf B, Smola A, Muller K. Nonlinear component analysis as kernel eigenvalue problem [J]. Neural Computation, 1998, 10 (5): 1299-1399.

[5] Yoo C K, Villez K, Lee I B, *et al.* Multivariate nonlinear statistical process control of a sequencing batch reactor [J]. Journal of Chemical Engineering of Japan, 2006, 39 (1): 43-51.

[6] 李元，马雨含，郭金玉. 基于动态多向局部离群因子的在线故障检测 [J]. 计算机应用研究, 2017, 34 (11): 3259-3261, 3266.

[7] 彭秀艳，柴艳有，满新江. 基于 PCA-KFCM 的船舶柴油机故障诊断 [J]. 控制工程, 2012, 19 (2): 311-315.

[8] 郭金玉，赵璐璐，李元. 基于统计特征的不等长间歇过程故障诊断研究 [J]. 计算机应用研究, 2014, 31 (1): 128-130.

[9] 郭飞，王成. 基于 LMP 和 KPCA 的人脸识别 [J]. 计算机工程, 2010, 36 (24): 183-186.

[10] Jenssen R. Kernel entropy component analysis [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2010, 32 (5): 847-860.

[11] 解亚萍，赵鹏，党伟明. 基于 KMC-KECA 的间歇发酵过程的故障诊断 [J]. 石油化工自动化, 2016, 52 (6): 21-26.

[12] 高学金，薛攀娜，齐咏生，等. 基于子时段 MPCA-SVM 的间歇过程在线故障诊 [J]. 计算机与应用化学, 2016, 33 (4): 465-471.

[13] 何青，褚东亮，毛新华，等. 基于 EEMD 和 MFFOA-SVM 滚动轴承故障诊断 [J]. 中国机械工程, 2016, 27 (9): 1191-1197.

[14] Tan Ying, Zhu Yuanchun. Fireworks algorithm for optimization [C]// Proc of International Conference on Swarm Intelligence. 2010: 355-364.

[15] Jenssen R, Eltoft T, GirolamiI M. Kernel maximum entropy data transformation and an enhanced spectral clustering algorithm [C]// Advances in Neural Information Processing Systems. 2007: 633-640.

[16] 于世飞，齐丙娟，谭红艳，等. 支持向量机理论与算法研究综述 [J]. 电子科技大学学报, 2011, 40 (1): 2-10.

[17] 谭营，郑少秋. 烟花算法研究进展 [J]. 智能系统学报, 2014, 9 (5): 515-528.

[18] Zhang Y, Zhang Y. Complex process monitoring using modified partial least squares method of independent component regression [J]. Chemometrics & Intelligent Laboratory Systems, 2009, 98 (2): 143-148.

[19] Yu J. Multiway Gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes [J]. Industrial & Engineering Chemistry Research, 2012, 51 (40): 13227-13237.

[20] 郑皓，熊伟丽，徐保国，等. 一种基于 LSSVM 的间歇过程在线故障诊断方法 [J]. 计算机与应用化学, 2017, 34 (1): 30-34.